Remarks/Arguments:

Because of the complex nature of the subject area, Applicants have addressed the rejection as follows. A status of the claims is provided under §1. A summary of the personal interview is provided under §2. Remarks related to paragraphs 3-6 of the Office Action are then provided under §3. Remarks related to paragraphs 8-16 of the Office Action are provided under §4. Conclusions are provided under §5. An appendix is provided to contrast an example of processing according to Pfister et al. and processing according to Applicants' invention.

1. Status of Claims

Claims 1, 3-13, 15, 17, 19, 21 and 22 are pending. Claims 1, 5, 9, 13, 15, 17, 19 and 21 have been rejected under 35 U.S.C. § 102(b) as being anticipated by Pfister et al. (WO 96/03741). Claim 9 has been rejected under 35 U.S.C. § 103(a) as being unpatentable over Pfister et al. in view of an Official Notice. Claims 3, 6-7, 10-11 and 22 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Pfister et al. in view of Abe et al. (U.S. Pat. No. 6,173,253). Claims 4, 8 and 12 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Pfister et al. in view of Abe et al. and further in view of Huang et al. (U.S. Pat. No. 5,829,000).

2. Personal Interview

Applicants acknowledge with thanks the courtesy shown to their representative by Examiner Vo during the personal interview of November 15, 2005. No agreement was reached during the course of the interview.

3. Remarks related to Paragraphs 3-6 of the Office Action

A. Summary of the Invention

The present invention is a method and apparatus for converting inputted speech to text. In Applicants' claim 1, an utterance is defined as being composed of a plurality of word-strings. Each word-string is defined to include one or more words. Candidates of word-strings are defined to consist of one or more words of the inputted utterance.

The method a) inputs the utterance. The method next b) performs speech recognition processing on one of the word-strings of the utterance to determine candidates of word-strings for that one word-string. The method c) displays the candidates. In step d) one of the displayed candidates is then selected by a user.

The method then performs speech recognition, displays and selects candidates, steps b)-d), for each successive word-string in the utterance until the end of the utterance is reached. Thus, speech recognition is performed on word-strings rather than the entire utterance. Furthermore, correction of speech recognition is performed during the speech recognition processing of the utterance rather than afterwards.

B. Argument

Claims 1, 5, 9, 13, 15, 17, 19 and 21 have been rejected under 35 U.S.C. § 102(b) as being anticipated by Pfister et al. (WO 96/03741). The rejection is respectfully traversed. It is respectfully submitted that these claims are patentable over the cited art for the reasons set forth below.

Claim 1 includes features neither disclosed nor suggested by the cited art, namely:

> ... inputting an utterance ... comprised of a plurality of word-strings ... each include one or more words ...
>
> ... determining candidates of word-strings ... by performing speech recognition processing on one of the plurality of word strings of the utterance ...
>
> ... displaying the candidates ...
>
> ... selecting one of the displayed candidates by a user ...
> ... said candidate determining step (b), said displaying step (c) and said selecting step (d) are repeated for each successive word-string in said utterance until an end of the utterance is reached ... (Emphasis added)

Pfister et al. disclose a speech transcription system that provides a training mode, a dictation mode and a display and editing mode, as discussed in §3C. Pfister et al. first perform phoneme recognition on the entire utterance in dictation mode, as discussed in §3D. Pfister et al. disclose a method of phoneme recognition processing

and that phoneme recognition processing is performed on the entire utterance, discussed in §3E. Pfister et al. then edits phonetic symbol string or identifies word boundaries within the recognized phonetic symbol string in display and editing mode after receiving the phonetic symbol string for the entire utterance, discussed in §3F.

Therefore, Pfister et al. first perform phoneme recognition on an utterance and then allow a user to correct the recognized phonetic symbol string or to select words corresponding to word boundary candidates within the phonetic symbol string, after phoneme recognition on the entire utterance is performed, using the display and editing mode. The user can edit and thus provide feedback to the system for the recognized phonetic symbol string representing the entire utterance within the display and editing mode.

Pfister et al. do not disclose or suggest Applicants' claimed features of "(a) determining candidates of word-strings ... by performing speech recognition processing on one of the plurality of word strings of the utterance ... (b) displaying the candidates ... (c) selecting one of the displayed candidates by a user ... step (b), ... step (c) and ... step (d) are repeated for each successive word-string in said utterance until an end of the utterance is reached ..." (emphasis added). More specifically, Applicants perform speech recognition on successive word-strings of an utterance using a user-selected candidate for a current word-string within the utterance. Thus, correction of misrecognition is performed during the recognition procedure of an utterance.

These features are neither disclosed nor suggested by Pfister et al. As discussed above, Pfister et al. correct misrecognition after recognition. Thus, Pfister et al. do not include all of the features of claim 1. Applicants' claimed features provide an advantage over Pfister et al. by requiring less storage capacity and less computation by performing speech recognition on word-strings rather than on an entire inputted utterance. Accordingly, allowance of claim 1 is respectfully requested.

C. Pfister et al. Provide a Training Mode, a Dictation Mode and a Display and Editing Mode

Pfister et al. disclose a speech transcription system that is based on phoneme recognition (p. 10, line 33). The system has a training mode, a dictation mode and a display and editing mode (p. 6, lines 8-11). Training mode is used to better identify the speaker's words when operating in dictation mode (p. 6, lines 18-23). Dictation mode generates a machine readable phonetic symbol string of an utterance according to phoneme recognition (p. 9, line 23- p. 19, line 35). Display and editing mode allows editing of the recognized phonetic symbol string (p. 21, line 10- 32) and further transcription of the phonetic symbol string into words according to word bound candidates (p. 22, line 15 p. 27, line 30).

D. Pfister et al. Perform Phoneme Recognition on Entire Utterance in Dictation Mode

In dictation mode, a user first dictates an utterance, such as "the sky is clear". Then spectral features are extracted from the utterance (p. 15, lines 31-33). Next, phoneme recognition is performed on the entire utterance using the spectral features and phoneme models (p. 15, line 35-p. 19, line 21). Then the system generates a machine readable phonetic symbol string to represent the utterance according to the phoneme recognition (p. 19, line 23-25).

E. Pfister et al. disclose a Method of Phoneme Recognition Processing and that Phoneme Recognition Processing is Performed on Entire Utterance

Phonemes represent the smallest contrastive sound unit in speech. They may represent types of vowels, consonants and diphthongs. For example, the sound of "r" in red, bring and round is a phoneme. Pfister et al. show common phonemes used in western, European languages in Table 1 (p. 16). Pfister et al. disclose a forward pass processing (p. 17, lines 1-4) and a backward labeling processing (p. 18, lines 18-22) in order to identify the most likely phoneme candidates for a speech segment. The processing continues "for the next and subsequent speech segments within the speech signal" (p. 18, lines 23-25). Thus, phoneme recognition is performed on the entire utterance.

F. Pfister et al. Edits Phonetic Symbol String or Identifies Word Boundaries within Recognized Phonetic Symbol String in Display and Editing Mode after Receiving Phonetic Symbol String for the Entire Utterance

Pfister et al. disclose that display and editing mode occurs once machine-readable phonetic symbols corresponding to the last spoken speech is received (p. 20, lines 30-33). In display and editing mode, the phonetic symbol string can be presented and edited by the user (p. 21, lines 16-32). Alternatively, the system can identify possible word boundaries within the phonetic symbol string (p. 22, lines 15-17) so that the phonetic symbol string can be transcribed into words. Words based on word boundary candidates are presented to a user. Once the user selects a word, the word boundary for the first word is set and processing continues within the recognized phonetic symbol string to find next word boundary candidates within the phonetic symbol string (p. 27, lines 24-30).

G. Remarks regarding claims 5, 9, 13, 15, 17, 19 and 21

Although not identical to claim 1, claim 5 includes features similar to claim 1 which are not disclosed or suggested in the cited art, namely, performing speech recognition on successive word-strings of an utterance using a user-selected candidate for a current word-string within the utterance. Accordingly, allowance of claim 5 is respectfully requested.

Claim 9 includes all of the features of claim 5 from which it depends. Accordingly, claim 9 is also patentable over the cited art.

Although not identical to claim 1, claim 13 includes features similar to claim 1 which are not disclosed or suggested in the cited art, namely, performing speech recognition on successive word-strings of an utterance using a user-selected candidate for a current word-string within the utterance. Accordingly, allowance of claim 13 is respectfully requested.

Claim 15 includes all of the features of claim 13 from which it depends. Accordingly, claim 15 is also patentable over the cited art.

Although not identical to claim 1, claim 17 includes features similar to claim 1 which are not disclosed or suggested in the cited art, namely, performing speech recognition on successive word-strings of an utterance using a user-selected candidate for a current word-string within the utterance. Accordingly, allowance of claim 17 is respectfully requested.

Claim 19 includes all of the features of claim 17 from which it depends. Accordingly, claim 19 is also patentable over the cited art.

Claim 21 includes all of the features of claim 1 from which it depends. Accordingly, claim 21 is also patentable over the cited art.

4. Remarks related to Rejection of Paragraphs 8-16 of the Office Action

Claim 9 has been rejected under 35 U.S.C. § 103(a) as being unpatentable over Pfister et al. in view of an Official Notice that cellular telephones having speech recognition capability is well known in the art. Claim 9, however, includes all of the features of claim 5 from which it depends. Accordingly, claim 9 is also patentable over the cited art.

Claims 3, 6-7, 10-11 and 22 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Pfister et al. in view of Abe et al. (U.S. Pat. No. 6,173,253). Claims 3 and 22, however, include all of the features of claim 1 from which they depend. Claims 6-7 and 10-11 include all of the features of claim 5 from which they depend. Abe et al. do not make up for the features lacking in Pfister et al. Accordingly, claims 3, 6-7, 10-11 and 22 are also patentable over the cited art.

Claims 4, 8 and 12 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Pfister et al. in view of Abe et al. and further in view of Huang et al. (U.S. Pat. No. 5,829,000). Claim 4, however, includes all of the features of claim 1 from which it depends. Claims 8 and 12 include all of the features of claim 5 from which they depend. Huang et al. do not make up for the features that are lacking in Pfister et al. and Abe et al. Accordingly, claims 4, 8 and 12 are also patentable over the cited art.

5. Conclusion

In view of the arguments set forth above, the above-identified application is in condition for allowance, which action is respectfully requested.

Appendix

A. Example of Utterance Transcription According to Pfister et al.

For example, assume that the utterance "the sky is clear" is processed. A separate sheet is enclosed to illustrate the process of phoneme recognition and subsequent word boundary determination. As shown in step 1, Pfister et al. will process the utterance by performing phoneme recognition on the entire utterance "the sky is clear." In step 2, the phoneme recognized utterance is converted to a phonetic symbol string. Note that the International Phonetic Alphabet (IPA) for English sounds is used in the example to illustrate phoneme symbols.

In step 3, the system determines word boundaries candidates to represent the first word within the phonetic symbol string. For the first word, a phonetic symbol substring is selected beginning at the first phoneme symbol. A first boundary, B1, may include only the first symbol. A second boundary, B2, includes the first and second symbol. A third boundary B3, includes the first-third symbols.

In step 4, the substrings SUB_{B1}, SUB_{B2} and SUB_{B3} are converted to word candidates, for example, "a", "the" and "this". In step 5, the word candidates are presented to a user ranked in an order of linguistic usage (p. 23, lines 9-12). The user, for example, selects "the" corresponding to word boundary B2.

After the user selects the first word, the word boundary B2 is used, in step 6, so that word boundary processing continues at the symbol following the selected B2 boundary (p. 26, lines 25-29). Candidates of word boundaries a subsequent word are selected from the remaining portion of the recognized phonetic symbol string. In step 7, the word boundary process is repeated (steps 3-7) to determine word boundaries within the remainder of the phonetic symbol string.

Thus, in Pfister et al., once a phonetic symbol string is recognized in dictation mode, word boundaries within the recognized phonetic string are determined afterwards, in display and edit mode. The recognized phonetic symbol string represents the entire utterance. Pfister et al. only determines word boundary

candidates on the recognized phonetic symbol string in order to transcribe the recognized phonetic symbol string into words.

B. Example of Utterance Transcription According to Applicants' Claim 1

By contrast, Applicants' claim 1 will process the sentence by first processing a word-string portion of the utterance, for example, "the sky" to determine candidates based on that portion of the spoken utterance. Next, candidates such as "the sky", "the pie", and "to buy" are displayed. Next, a user selects a best match, for example, "the sky". Next, steps of speech recognition processing to determine candidates, display of candidates, and user-selection of candidates are repeated on a successive word-string, for example "is". The process is repeated until the end of the utterance "clear" is reached. Thus, a user selects from among displayed word-strings as determined by determining candidates step of Applicants' method. Furthermore, speech recognition processing is performed on a word-string consisting of one or more words. Furthermore, a user-selection is performed at each successive word-string until the end of the utterance is reached.

Respectfully submitted,

RatnerPrestia

Lawrence E. Ashery, Reg. No. 34,515
Attorney for Applicants

DG/dmw/fp

Dated: January 10, 2006

P.O. Box 980
Valley Forge, PA 19482
(610) 407-0700

The Commissioner for Patents is hereby authorized to charge payment to Deposit Account No. 18-0350 of any fees associated with this communication.

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail, with sufficient postage, in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on:

January 10, 2006

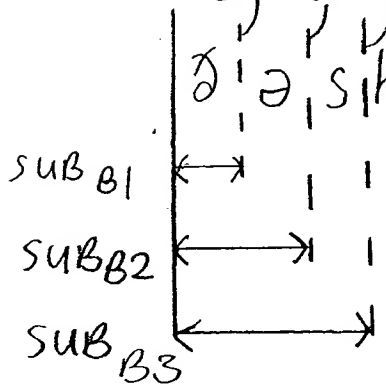
FP_I:\MAT\8201US\AMEND03.DOC

1. "the sky is clear"

↓ phoneme recognition

2. "ðəskaIIzklIə"

3.



4.

B1 ðə → WORD B1 (a)

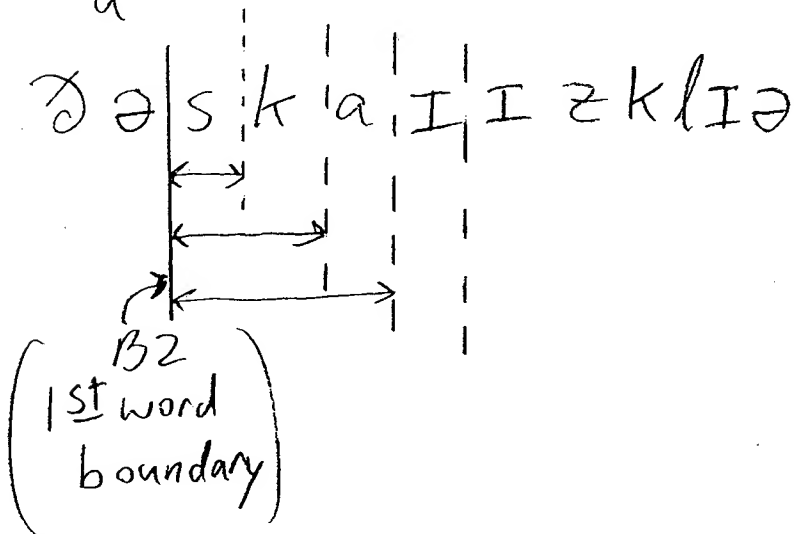
B2 ðəs → WORD B2 (the)

B3 ðəs → WORD B3 (this)

5.

this
the ← select B2
a

6.



7. Repeat candidates of word boundaries (steps 3-6) on subsequent portion phoneme symbol STRING